

---

# **Tracking student movement using NHS Central Register (NHSCR) and Higher Education Statistics Agency (HESA) data**

**Alternative Sources Branch  
National Records of Scotland  
customer@gro-scotland.gsi.gov.uk**

---

# A story in three parts

- A hard to track group (students who leave home to study)
- A topical method for studying them (probabilistic record linkage)
- New results (information not otherwise available)

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future

---

# A hard to track group

- Students who leave home to study should register with a GP in the place where they are studying (which in the present case is England or Wales)
- If they do, their NHS Central Register (NHSCR) records will be updated to show the move
- But how often do they register (and hence how often do they leave their NHSCR records in error?)

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future

---

# Two data sets needed

- NHSCR extract lists the students with their (sometimes wrong) postcodes
- The Higher Education Statistic Agency (HESA) database of Scottish-domiciled students studying in England or Wales lists the term time postcode
- To compare the HESA post code (which we assume to be correct) with the NHSCR post code requires linking the two data sets

---

# The HESA data set

- 47,549 Scottish-domiciled students in 2007/8 and/or 2008/9 academic sessions
- But it turned out that it these were students studying *at universities located* in England or Wales - many records were of Open University (OU) students on distance learning courses etc
- No unique data trace for students located in England or Wales
- Records were removed if term time post code was Scottish (but not if it was missing)
- or if the date of birth was before 1984 or after 1990 (i.e. they had to be between 18 and 26 in 2008)

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future

---

# The NHSCR data set

- 8,704,299 people on the extract in July 2010
- Sample was again limited to dates of birth from 1984 to 1990 (ie from 18 to 26 in 2008)
- Post code valid on 1 September 2008 was identified

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future

---

# The linkage procedure (1)

- 665,000 NHSCR records (7.6% of the total) and 6,909 HESA records (14.5% of the total)
- Two packages used were Link Plus (US Centre for Disease Control) and Rec Link (US Census Bureau)
- Of the HESA records, 6,534 (94.6%) were confidently linked to the NHSCR file
- What does “confident” mean?

---

# The linkage procedure (2)

- Two types of errors
- Accepting false links (do not in fact refer to the same person)
- Not accepting true links (do in fact refer to the same person)
- Unavoidable tension between these two
- Balance struck by the relative costs of the two error types



---

# The linkage procedure (3)

- First Middle Last Birthdate Homepostcode Gen
- ANNE HELEN ROBERTSON 19880624 PA78 6TA F
- ANNE HELEN HENDERSON 19880624 PA78 6SB F
- OLIVIA KATE ABRAN 19870211 EH4 1QX F
- OLIVIA ABRAM 19870211 G13 8NN F
- JOHN SMITH 19880503 KY14 3LR M
- JOHNSMITH 19880503 KY14 3LR M
- HO CHI MINH 18900519 DG2 8RL M
- CHI-MINH HO 18900519 DG2 8AN M

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future

---

# The linkage result

- Of the HESA records, 6,534 (94.6%) were confidently linked to the NHSCR file
- The 375 who could not be linked were more likely to be of non-GB nationality (12% vs 4%) and non-white ethnicity (19% vs 10%)
- Using only linked records for students following first degrees and for whom complete data is available leaves 3,893 records.

---

3,893 first degree students domiciled in Scotland but studying in  
England or Wales in 2007-9

Location of health board registration at end of academic year by year of  
study

**Year of study**

<b>Region</b>	<b>1st</b>	<b>2nd</b>	<b>3rd</b>	<b>4th and beyond</b>
<b>Scotland</b>	<b>27%</b>	<b>19%</b>	<b>18%</b>	<b>14%</b>
<b>England / Wales</b>	<b>71%</b>	<b>79%</b>	<b>81%</b>	<b>85%</b>
<b>Other</b>	<b>2%</b>	<b>2%</b>	<b>1%</b>	<b>1%</b>
<b>Total</b>	<b>1,279</b>	<b>1,157</b>	<b>987</b>	<b>470</b>

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future

---

# Conclusions

- Probabilistic record linkage offers the best way of integrating data from disparate sources where there is no shared unique identifier
- It raises a wide range of issues from highly technical statistics to legislation and public perception
- But it is not without its difficulties
- Clerical review is time consuming (and therefore expensive), repetitive and requires constant concentration
- But without it error rates increase.
- Even so, it is the future for data integration

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future

- 
- Opinions?
  - Comments?
  - Questions?
  - Reactions?

**NATIONAL RECORDS OF SCOTLAND**

preserving the past; recording the present; informing the future